CSCI 4144, Winter 2014

Assignment 2: DW/OLAP & DM Exercises Using SQL Server2012

Issue: Jan. 21
Due: Feb. 04
TA: Mayank Malhotra <malhotra@cs.dal.ca>
Tutorial: 01/23, 1:00-2:00 pm; 01/24, 5:00-6:00 pm, in Teaching Lab 2
Help Hours: 01/28, 1:00-2:00 pm; 01/30, 1:00-2:00 pm; 01/31, 3:30-4:30 pm, in CS L-Center

---

1. Objectives:

1) To learn how to use a commercial package to do DW/OLAP and DM work.
2) To gain an in-depth understanding of DW and OLAP concepts via developing.
3) To get some insights on what kinds of knowledge can be mined and presented by applying available data mining modules and visualization tools.

2. Data DW/OLAP Exercises:

1) Select a data set to work with:

- You can either use the given database (DB), or a DB of your own, or a DB from a data repository (e.g. http://www.ics.uci.edu/~mlearn/MLRepository.html), which is suitable for this exercise.
- Write up an application scenario (i.e. describe the application and the requirements, etc.) based on the chosen DB.
- **Note:** The DB to be used must have hierarchical concepts for some dimensions (i.e. attributes of the DB).

2) If the selected DB is not in relational format, you need to convert it into a relational one (the DW software can only recognize relational tables). You can use the DBMS tool to achieve it (i.e., get data prepared).

* You may perform normalization to arrive at the relational model. Practically, you could utmost import this data into a single table. Then, you will have to manually "normalize" that table into some other tables, so that you end up with many tables having parent-child relationship. This parent-child relationship is established using, referential integrity or foreign key references using SQL server DDL statements. The most common normal forms out there are: 1NF, 2NF, 3NF, Boyce-codd normal form and 4NF. Most RDBMS are in 3NF.

3) Use Snowflake schema to design a data cube and then materialize it by applying the SQL Server's DW tools.

4) Design a set of typical OLAP queries (at least: 4, i.e. one for each) for the operators: roll-up, drill-down, slice, and dice. Choose a proper visualization tool to present each result. To include screenshots of each result, double-click your cube, which opens up the "your-cube-name.cube [Design]" window under which you need to select the "Browser" tab. Include the screenshots of your results that is, displayed on this "Browser" tab.

5) Write a report including the following sections:

(a) Task Description:
Present the business scenario (i.e., the application and the requirements), and the DB (provide an instance of the database, and the reference source, etc.).

E.g. See http://technet.microsoft.com/en-us/library/ms124825(v=SQL.100).aspx
Check out the links given in the above URL under "In This Section".

(b) Data Cube Design:
Provide a DW schema diagram and an instance fragment of each table.

(c) Application Queries:
Each query is an ad hoc business question (in English) about some analytical information on the subject, which is then translated into OLAP operations, i.e. by choosing dimensions/concepts/values for producing a report (i.e. the retrieved information is represented in an analytical screen (table) form).

(d) Summary:
Provide a summary of your work & observations on the application (i.e. the DB), the DW and the software tools (comments & recommendations, etc).

3. Data Mining Exercise:

1) Apply the available DM modules, i.e. classification (Microsoft DT) and clustering (Microsoft Clustering), to the selected data sets for mining patterns hidden in the data.

2) Write a DM report to present each of the DM applications.
   (a) A brief application scenario. Check the link shown in DW section.
   (b) A brief description of the procedure you applied for mining a result.
   (c) Result presentation (use a proper visualization tool) and possible interpretation of the result.
   (d) Any other comments.

4. Submit your Ass2 report electronically:

1) Create a directory from your home directory on bluenose and name it as assign2.

2) Place your Ass2 document in the directory.

3) Submit it from your home directory by the command line: submit assign2
(you will receive a message when you hit the Enter key).

5. References:

- Doc/SQL-Server2012Guide.pdf
- Doc/Ass2Tutorial.pdf
- Creating Your First Multidimensional Cube:
http://www.packtpub.com/sites/default/files/9781849689502_Chapter_03.pdf

*** Plagiarism and Intellectual Honesty:** (http://plagiarism.dal.ca)
Dalhousie University defines "plagiarism as the presentation of the work of another author in such a way as to give one's reader reason to think it to be one's own." Plagiarism is considered a serious academic offense which may lead to loss of credit, suspension or expulsion from the University, or even the revocation of a degree.